Místo a role zrcadlových neuronů ve vtělených kognitivních systémech

Wiedermann, Jiří
2003

Institute of Computer Science
Academy of Sciences of the Czech Republic

Místo a role zrcadlových neuronů ve vtělených kognitivních systémech

Jiří Wiedermann

Technical report No. 888

březen 2003

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: (+4202) 6605 3520, fax: (+420) 286 585 789,
e-mail:jiri.wiedermann@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Místo a role zrcadlových neuronů ve vtělených kognitivních systémech

Jiří Wiedermann[1]

Technical report No. 888

březen 2003

Abstrakt:

Práce se zabývá otázkou osvětlení algoritmických mechanizmů, stojících v pozadí rozvoje mentálních schopností přirozených anebo umělých kognitivních systémů. Východiskem úvah je formální model kognitivního agenta, jehož návrh je motivován moderními přístupy ke vtělené kognici. Uvažujeme efektivní realizaci takového agenta pomocí neuronových sítí, ve kterých důležitou roli hrají zrcadlové neurony. V našem modelu je jejich hlavním posláním koordinace senzomotorické informace a její doplňování v případech, kdy některá složka informace chybí. Tyto případy odpovídají imitačnímu učení, empatii a myšlení. Dále poukážeme na význam propriocepce při formování subjektivních prožitků. Ukážeme, že zrcadlové neurony a doplňovací mechanizmy informací tvoří základ, pomocí kterého je možné dále budovat hypotézy o rozvoji komunikačních schopností vedoucích od posunkového jazyka až po rozvoj myšlení. V těchto hypotézách hraje hlavní roli asociování motorické informace s odpovídající percepční a propriocepční informací a postupné utlumování odpovídajících lokomočních akcí. Práce přináší nové výsledky potvrzující oprávněnost nadějí, vkládaných do objevu zrcadlových neuronů.

Keywords:
vtělená kognice, autonomní agent, zrcadlové neurony, neuronové sítě, evoluční rozvoj kognitivních schopností

[1] Tato práce vznikla s částečnou podporou grantu GA ČR č. 201/02/1456

MÍSTO A ROLE ZRCADLOVÝCH NEURONŮ VE VTĚLENÝCH KOGNITIVNÍCH SYSTÉMECH

Jiří Wiedermann[1]
Ústav informatiky AV ČR
Pod vodárenskou věží 2, 182 07 Praha 8
Česká republika
jiri.wiedermann@cs.cas.cz

Motto: *„The discovery of mirror neurons in the frontal lobes of monkeys, and their potential relevance to human brain evolution ... is the single most important "unreported" (or at least, unpublicized) story of the decade. I predict that mirror neurons will do for psychology what DNA did for biology: they will provide a unifying framework and help explain a host of mental abilities that have hitherto remained mysterious and inaccessible to experiments. "*

V.S. Ramachandran [1]

Abstrakt: Práce se zabývá otázkou osvětlení algoritmických mechanizmů, stojících v pozadí rozvoje mentálních schopností přirozených anebo umělých kognitivních systémů. Východiskem úvah je formální model kognitivního agenta, jehož návrh je motivován moderními přístupy ke vtělené kognici. Uvažujeme efektivní realizaci takového agenta pomocí neuronových sítí, ve kterých důležitou roli hrají zrcadlové neurony. V našem modelu je jejich hlavním posláním koordinace senzomotorické informace a její doplňování v případech, kdy některá složka informace chybí. Tyto případy odpovídají imitačnímu učení, empatii a myšlení. Dále poukážeme na význam propriocepce při formování subjektivních prožitků. Ukážeme, že zrcadlové neurony a doplňovací mechanizmy informací tvoří základ, pomocí kterého je možné dále budovat hypotézy o rozvoji komunikačních schopností vedoucích od posunkového jazyka až po rozvoj myšlení. V těchto hypotézách hraje hlavní roli asociování motorické informace s odpovídající percepční a propriocepční informací a postupné utlumování odpovídajících lokomočních akcí. Práce přináší nové výsledky potvrzující oprávněnost nadějí, vkládaných do objevu zrcadlových neuronů.

Klíčová slova: vtělená kognice, autonomní agent, zrcadlové neurony, neuronové sítě, evoluční rozvoj kognitivních schopností

[1] Tato práce vznikla s částečnou podporou grantu GA ČR č. 201/02/1456

1. ÚVOD

Na přelomu osmdesátých a devadesátých let 20. století objevil v mozku opic Giacomo Rizzolatti z Parmské univerzity a v řadě publikací (viz např. [9]) popsal nový druh neuronů. Tyto tzv. zrcadlové neurony jsou aktivní, když jejich majitel provádí jisté vysoce specializované pohyby rukou, např. tahání, tlačení, škubání, uchopení, utržení ořechu a vložení do úst, apod. Jak píše Ramachandran [1], nelze se ubránit dojmu, že se jedná o tytéž neurony, které vydávají příslušné pohybové příkazy pro svaly. Co je však zvláštní a nečekané je skutečnost, že stejná skupina zrcadlových neuronů, která byla aktivní v případě realizace nějakého příkazu, je aktivní i v případě, když opice pozoruje jinou opici, která vykonává stejnou činnost! V práci [5] můžeme dokonce nalézt následující „definici“: zrcadlový neuron je neuron, který je aktivní jak v případě, kdy subjekt vykonává nějakou akci, tak i v případě, kdy subjekt pozoruje stejnou akci vykonávanou jiným subjektem. To lze interpretovat i tak, že u opic jsou zrcadlové neurony jakýmsi mechanismem pro „čtení myšlenek“ (mind reading) jiných opic. Další badatelé tyto fakty zobecnili i na další primáty a z tohoto faktu vyvodili a vyvozují další závěry, týkající se významu tohoto objevu pro porozumění úmyslů jiných osob, empatii, imitačnímu učení, a dokonce evoluci lidské řeči [1], [5]. Současně, ale nezávisle na objevu zrcadlových neuronů, se v kognitivních vědách začalo prosazovat nové paradigma tzv. vtělené kognice (embodied cognition – viz např. [7]). V porovnání s klasickým přístupem ke kognici, reprezentovaným klasickou umělou inteligencí, zdůrazňuje vtělená kognice význam interakce systému s prostředím, jeho vtělenost (inkarnaci – systém musí mít „tělo“) a situovanost (systém musí být v každém okamžiku „v obraze“ dění). Tento přístup znamená významný posun jak na úrovni explanace a pochopení kognitivních mechanizmů, tak i – a možná zejména -- na úrovni realizační. Zdá se však, že místo a role zrcadlových neuronů se zatím v kontextu vtělené kognice nijak zvlášť nestudovala.

Cílem této práce, která má předběžný, explorační charakter, je ukázat možné místo a možnou roli zrcadlových neuronů v rozvoji mentálních schopností v kontextu vtělené kognice. Pro tyto účely využijeme klasické formální prostředky -- modelování vtěleného kognitivního agenta pomocí konečných automatů a neuronových sítí. Ukážeme, že vlastnosti našeho modelu, které zobecňují pozorované vlastnosti zrcadlových neuronů, jsou stále v souladu s empirickými poznatky o nich a na základě dalších (teď již neformálních) úvah o rozvoji kognitivních schopností modelu naznačíme hypotézy, které potvrzují klíčovou roli zrcadlových neuronů pro pochopení mechanizmů kognice.

Ve 2. části práce navrhneme formální model kognitivního agenta a stručně podiskutujeme motivace předloženého návrhu. Tento agent bude modelován pomocí konečného transduceru, který je situován ve svém prostředí pomocí tzv. percepčně--motorických orgánů. Dále, ve 3. části na příkladu speciální kognitivního agenta, známého z odborné literatury pod názvem houbožer (fungus eater – viz např. [7]), ukážeme mechanizmus, pomocí kterého lze modelovat nejjednodušší případ učení pomoci imitace – opakování pozorovaného pohybu. Základem tohoto mechanizmu je doplňováni motorické informace k pozorované senzorické informaci. Ve 4. části ukážeme realizaci doplňovacího mechanizmu multimodální informace pomocí zrcadlových neuronů, resp. neuronových sítí a vysvětlíme důvody, proč lze z hlediska agenta takovou síť považovat za základ modelu jím poznatelného světa. V 5. části

představíme hypotézu o evolučním rozvoji mentálních schopností kognitivních agentů, která důsledně vychází z předchozích představ o koordinaci a doplňování mutimodální informace pomocí zrcadlových neuronů. V závěru práce shrneme dosažené výsledky a jejich význam v kontextu snah o výpočetní modelování mysli.

2. KONEČNÝ KOGNITIVNÍ AGENT

Centrálním předpokladem moderní kognitivní vědy je hypotéza, že kognice není nic jiného než jisté specifické zpracování informací, souvisejících s kognicí, tj. specifický druh výpočtů nad specifickými daty. Jako metafora mechanizmu vlastní kognice se proto často používá Turingův stroj, který je chápán jako čítankový příklad universálního mechanismu pro realizaci výpočtů.

Tato metafora má kořeny pravděpodobně již v úvahách samotného Turinga, který navrhl svůj model jako vysoce zjednodušenou analogii práce matematika. Turingův stroj je z jedné strany velmi hrubým modelem kognice, protože (kromě jiného) není dostatečně strukturován, takže v architektuře Turingova stroje je těžké nalézt analogii jednotlivých komponentů kognitivních orgánů živých organismů. Přesněji řečeno, Turingovy stroje již ve svém návrhu „zakrývají" jeden důležitá aspekt kognice -- a sice zpětnou vazbu mezi motorickými akcemi subjektu a jejich následnou percepcí týmž subjektem pomocí vizuálního, akustického či jiného mechanizmu. Pokud v Turingově stroji motorickým akcím odpovídá pohyb hlavy Turingova stroje a percepci čtení symbolu z pásky, tak žádná zpětná vazba není potřebná, protože hlava Turingova stroje přesně vykoná povely řídící jednotky. Nemůže dojít k žádné „nepřesné" či chybné realizaci instrukcí. Proto originální model Turingova stroje vystačí s jedinou hlavou, která obstarává jak bezchybnou realizaci instrukcí, tak i bezchybné čtení. Turingův stroj je z výpočetního hlediska také přehnaně efektivním modelem, protože na své potenciálně nekonečné pásce si může zapamatovat libovolné množství dat (což zřejmě žádný konečný živý organismus nemůže). I z hlediska spolehlivosti je Turingův stroj je příliš idealizovaným modelem, protože nepočítá s nemožností přesné realizace svých instrukcí a vstupně-výstupních operací z důvodů nepředvídatelných vnějších okolností. Poslední výhrada – běžný Turingův stroj nemá možnost komunikace s jinými stroji.

Představme si však situaci, kdyby zapisovací hlava z různých důvodů nemohla vždy přesně realizovat instrukce řídící jednotky, tj. někdy by zapsala jiný symbol, anebo by vykonala pohyb jiným směrem, než bylo přikazováno. Jak by musel vypadat Turingův stroj, který by „fungoval" i v takové situaci? Řešením by zřejmě byla další hlava, řízená řídící jednotkou, která by „pozorovala" akce původní hlavy a „hlásila" by řídící jednotce, co vidí. Řídící jednotka by na základě porovnání této informace s instrukcí, kterou vyslala, mohla odhalit nesprávné vykonání své instrukce a zabezpečit nápravu. Přitom předpokládáme, že „vidící" hlava podává vždy pravdivé informace a že řídící jednotka pracuje bezchybně

Na základě analogie se živými tvory se nám z hlediska z hlediska výpočetní kognice vynořuje představa podstatně odlišného modelu než je klasický Turingův stroj. Součástí tohoto modelu

by vůbec nebyla (nekonečná) paměťová páska, a byla by v něm zřetelně odlišena „mysl“ (řídící jednotka) od „těla“, které by se skládalo z řady různých, ale spolupracujících „orgánů“ pro motoriku a percepci. Dále, tento model by nerealizoval jednorázové konečné výpočty, tak jako klasický konečný automat, ale by neustále interagoval se svým prostředím. V dalším si trochu formálněji popíšeme model tzv. kognitivního konečného agenta, který nám umožní přesnější diskusi kognitivních schopností a realizace takových zařízení.

Definice: Kognitivní konečný agent (KKA) je interaktivní konečně—stavový transducer (výpočetní zařízení, transformující proudy vstupních dat na proudy výstupních dat), které se skládá

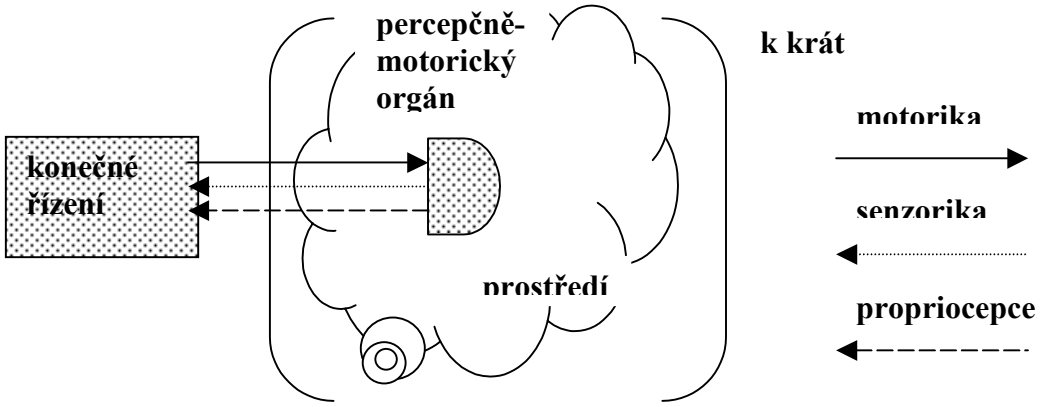
- z konečně stavové řídící jednotky
- minimálně z jednoho percepčně-motorického orgánu (PMO) .

Jak naznačuje její název, PMO je kombinací motorického a percepčního zařízení (podobně jako hlava Turingova stroje), tzn., že dostává od řídící jednotky motorické instrukce (instrukce pro pohyb) a současně posílá do řídící jednotky percepční informace, získané prostřednictvím čidel, kterými je zařízení vybaveno. Tyto percepční informace jsou dvojího druhu: senzorické informace, přicházející od čidel, které reagují na vlastnosti vnějšího světa (světelné, zvukové, sluchové, hmatové, tepelné, elektromagnetické podněty apod.) a tzv. propriocepční informace, přicházející od čidel, umístněných uvnitř systému a hlásících informace např. o svalovém napětí jednotlivých svalových skupin, pocity hladu, bolesti, úbytek energetických zdrojů, apod. Propriocepční informace tedy odpovídají jakýmsi „vnitřním pocitům či prožitkům“ při vykonávaní jednotlivých motorických činností. Všem třem druhům informací – motorickým senzorickým a propriocepčním – budeme říkat multimodální informace.

Příkladem percepčně-motorického orgánu je oko. Motorickým instrukcím odpovídají instrukce pro okolumotoriku (natočení oka, zaostření čočky, nastavení clony), senzorickým informacím odpovídá obrazová informace ze sítnice a propriocepční informace pochází od vnitřních sensorů oka, vypovídajícím o svalovém úsilí, nutném pro současné „nastavení“ ovládacích prvků oka. Podobně i např. ruka je příkladem percepčně-motorického orgánu: motorické informace jsou signály pro různé svalové skupiny ruky, senzorické informace jsou hmatové signály a opět jsou zde i propriocepční signály od jednotlivých svalových skupin. V robotických systémech je percpečně-motorickým orgánem i např. motorem poháněné kolo. Motorická instrukce je např. typu „pomaly vpřed“, percepční informací může být „kolo se neotáčí“.

Činnost KKA je definovaná pomocí přechodové funkce, která na základě současného stavu, současně vydaných instrukcí pro motorickou část PMO a současných informací od senzorů PMO přiřazuje nový stav a novou množinu instrukcí pro motorickou část PMO. Formálně, máme-li KKA s počtem k PMO a označíme-li Q konečnou množinu stavů řídící jednotky, M konečnou množinu instrukcí pro motorickou část PMO, S konečnou množinu informací od vnějších senzorů PMO a P množinu propriocepční informací od vnitřních sensorů, tak přechodová funkce má tvar $Q\times(M\times S\times P)^k \rightarrow Q\times M^k$. O množinách M a S budeme

předpokládat, že obsahují symboly, odpovídající situaci „žádný motorický signál nebyl vyslán“, resp. „žádná senzorická informace nebyla obdržena“. Propriocepční informace je vždy součástí percepce. Schéma KKA je na obrázku č. 1. Poznamenejme, že už i KKA odpovídající jednoduchým živým organizmům, jakým jsou např. jednobuněčné organizmy, mají řádově stovky PMO (viz např. [6]) a tisíce stavů, které jsou rozděleny na podmnožiny, odpovídajícím jistým kontextům (situacím), které KKA zjišťuje prostřednictvím svých senzorů a ve kterých se KKA může nacházet. Na KKA s velkým počtem stavů můžeme aplikovat principy subsumpce (viz. např. [2], [7]), protože v mnoha případech mohou vykonávat jednotlivé SMO svou činnost nezávisle na činnosti ostatních SMO (např. pohyb nemusí být potřebné koordinovat s čichovými vjemy apod.).



Obr.1: Konečný kognitivní agent

KKA se může pomocí PMO (např. kol, nohou) specializovaných na pohyb přesouvat v prostředí a pomocí dalších PMO (např. chapadel, rukou, manipulátorů apod.) může prostředí modifikovat; prostředí samotné však není předmětem našeho modelování. Informace o prostředí dostává KKA prostřednictvím některých (ne nutně všech) svých PMO (např. kamer). Každý PMO má tedy kromě motorické části odpovídající percepční část; její propriocepční složka obstarává *vnitřní zpětnou vazbu* od motorické jednotky. Příkladem může sloužit mechanické rameno, které dostává instrukce pro pohyb a vrací informace od pohonných mechanizmů, vypovídající o (úspěšnosti) provedení instrukcí. Navíc, některé PMO mohou poskytovat zpětnou vazbu jinému PMO. Typickým příkladem toho je systém oko-ruka. Tento systém dohromady má dvě motorické jednotky, jednu pro řízení pohybu ruky, druhou pro řízení pohybu oka. Percepční jednotky jsou také dvojí: jedna vrací informace od pohonných mechanismů ruky, druhá informace z oční sítnice. Dále zde máme propriocepci od obou orgánů. V případě, že „oko“ pozoruje svou vlastní „ruku“, říkáme, že dochází ke *vnější zpětné vazbě* mezi motorickou a vizuální jednotkou.

Oba druhy zpětných vazeb -- vnější i vnitřní – jsou důležité pro tzv. sebekontrolu KKA. Uvažujme pro jednoduchost KKA vybavený systémem oko--ruka a předpokládejme, že

automat je naprogramován tak, aby oko vždy sledovalo svou ruku. Přechodová funkce KKA by měla být taková, aby s instrukcemi pro každý pohyb ruky i oka byla asociována pomocí vnitřní zpětné vazby příslušná propriocepce, a pomocí vnější zpětné vazby mezi rukou a okem také koordinace jejich vzájemných pohybů. Takový režim práce KKA, kdy pohyby oka a ruky jsou koordinovány a propriocepce odpovídá takto prováděným pohybům celého systému, budeme nazývat *normálním režimem*. Je zřejmé, že jakmile nastane odchylka od normálního režimu, tak tato skutečnost bude indikována nesouhlasem asociovaných motorických a vnitřních či vnějších percepčních informací, které odpovídají normálnímu režimu. To znamená, že pokud je přechodová funkce KKA správně vytvořena, KKA má mechanizmus, jak tyto odchylky odhalit a dokonce může na danou nezvyklou situaci reagovat nápravními opatřeními. Tento mechanizmus budeme v dalším nazývat autokontrolním mechanismem a jeho realizací se budeme zabývat v další části.

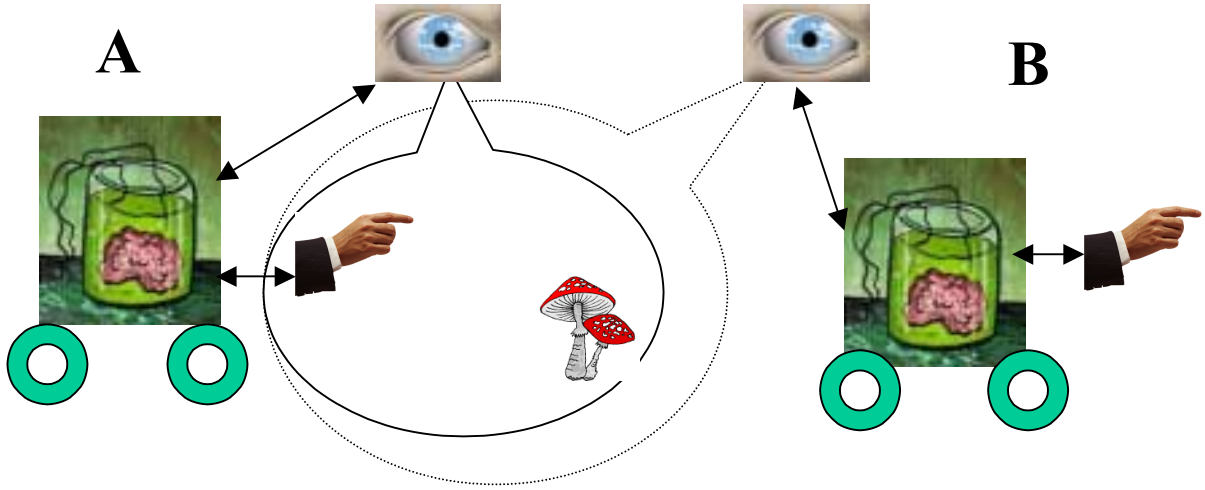
Poměrně lehce lze dokázat (a zde se právě hodí formalizovaný pojem KKA), že vhodně vybavený KKA ve vhodném prostředí může simulovat klasický Turingův stroj. V takovém případě potřebuje KKA dva různé PMO, které se pohybují synchronně a jeden se specializuje na realizaci zapisovací hlavy (propriocepce není využita) a druhý na realizaci čtecí hlavy Turingova stroje. Jako páska slouží vhodné prostředí, ve kterém se KKA lineárně pohybuje pomocí vhodného lokomočního mechanismu a zaznamenává do něj potřebné znaky, anebo je „čte“. Opačná simulace je možná pouze ve speciálních případech -- např. pro právě popsaný KKA, avšak pouze tehdy, kdy tento pracuje „bezchybně“, tj. spolehlivě vykonává příkazy řídící jednotky, takže percepční jednotky nikdy nehlásí chybnou realizaci těchto příkazů. Klasický Turingův stroj totiž nemůže odhalit situaci, kdy jeho hlava v důsledku nespolehlivosti či poruchy realizuje jiný pohyb než ten, který je určen přechodovou funkcí pro daný kontext. KKA je v tomto smyslu obecnějším zařízením než klasický Turingův strojem.

3. MECHANIZMUS AUTOKONTROLY A DOPLŇOVÁNÍ MULITIMODÁLNÍ INFORMACE

Úlohou mechanizmu autokontroly je realizovat prověrku správné realizace motorických příkazů v normálním režimu, tj. formálně tento mechanizmus kontroluje „správné“ hodnoty argumentu přechodové funkce tvaru $Q\times(M\times S\times P)^k$. Jinými slovy, tento mechanizmus zjišťuje, jestli hodnoty prvků množin Q, M, S a P odpovídají hodnotám nabývaným v normálním režimu. Navíc, tento mechanizmus musí „zafungovat“ i v případě, kdy některé z odpovídajících hodnot nejsou z různých příčin k dispozici, anebo nesouhlasí s hodnotami v normálním režimu. Podstatu uvedeného jevu si vysvětlíme na příkladě imitačního učení. Za tím účelem budeme modelovat situaci popsanou v úvodu práce, ve které byly odhaleny schopnosti zrcadlových neuronů, pomocí našeho modelu KKA, odpovídajícímu systému oko--ruka. K tomu zřejmě potřebuje dva stejné KKA, A a B. Předpokládejme, že oba automaty jsou stejně naprogramovány, tj. stejné instrukce realizují stejným způsobem a tomu odpovídá i stejná vnitřní či vnější zpětná vazba. Dále přepokládejme, že přechodová funkce obou transducerů odpovídá normálnímu režimu. Představme si nyní, že agenti A a B se dostanou do

takové vzájemném pozice, ve které oko agenta B „vidí“ rameno agenta A ze stejné perspektivy, z jaké předtím vidělo své vlastní rameno. Tato situace je naznačena na obr. č. 2 pro případ tzv. houbožerů – jednoduchých referenčních KKA, tradičně uvažovaných v robotické literatuře [7].

Pro stroj B nastane zvláštní situace: jeho oko hlásí pohyby, které neodpovídají normálnímu režimu. Tzn., že pozorovaný pohyb („vidím rameno, které trhá houbu“) obecně nesouhlasí s vydanou instrukcí pro pohyb ruky stroje B a pro tuto situaci přechodová funkce stroje B není definovaná. Mechanizmus autokontroly stroje B tuto situaci odhalí. Doplňme nyní přechodovou funkci stroje B tak, aby v takovém případě přešel agent do zvláštního, tzv. pozorovacího režimu (pro který je charakteristická vybraná podmnožina stavů), který znamená „pozoruji pohyb, pro který jsem nevydal příkaz“. Dále můžeme v tomto režimu „opravit“ či „doplnit“ přechodovou funkci tak, aby se pozorování agenta B doplnilo tou motorickou instrukcí („utrhni houbu“), která vedla agenta A k vykonání jeho pohybu, a také odpovídající propriocepční informací, a to vše stále v pozorovacím režimu. Oba tyto doplňky odpovídají tomu příkazu a té propriocepci, kterou má stroj B v normálním režimu, když pozoruje své vlastní rameno (jak trhá houbu).



Obr. 2: Základ imitačního učení: Agent B přejde do stavu „pozoruji pohyb, pro který mi nesouhlasí vydaný motorický příkaz s pozorováním“; následně agent B doplní motorický příkaz tak, aby souhlasil s pozorováním

Zde někde můžeme hledat základ mechanismu imitačního učení jako schopnosti přiřadit odpovídající instrukci pozorované akci (a tedy v případě potřeby ji zopakovat), a také základ empatie, tj. schopnosti doplnit i příslušné vnitřní pocity (propriocepci) k pozorované akci. Právě popsané jednoduché imitační učení jednoho pohybu může být základem komunikace – agent B totiž může vykonat nějaký pohyb – říkejme mu *posunek* – za účelem „vyslání signálu“ agentu A, který ho pozoruje. Pokud je tento posunek „vysílán“ vždy ve stejném

kontextu (čemu může odpovídat vybraný stav či třída stavů), tak agent A se může časem tento kontext naučit doplňovat pomocí mechanismu, kterého základ jsme právě popsali. Imitační učení a komunikace je tedy složitější záležitost než pouhá schopnost opakovat pozorované pohyby – v praxi jde většinou o opakování neznámé sekvence pohybů, a také o naučení se kontextu, ve kterém je vhodné naučenou sekvenci použít. Jakým mechanismem to lze zabezpečit ukážeme v 5. části.

V obecném případě se může stát, že chybí jakákoliv ze tří součástí multimodální informace. nejvýznačnější situace zachycuje tabulka na obr. č. 3. Zde je znázorněno, v jakých situacích se doplňuje chybějící informace, resp. se „nahrazuje“ nekonzistentní informace. V obou případech je výsledkem multimodální informace konzistentní s normálním režimem. Informace o tom, jaký typ „rekonstrukce“ multimodální informace byl proveden, je zachycena v příslušných stavech KKA, takže agent může s touto informací pracovat.

Senzorika	Propriocepce	Motorika	Doplní se	Režim
ano	ano	ano	nic	normální
ano	ne	ne	propriocepce a motorika	pozorování (napodobení, empatie, znaková komunikace)
ne	ano	ano	senzorika	jednání „naslepo“
ne	ne	ano	vše	myšlení

Obr. 3: Doplňováni chybějící multimodální informace

Z předchozích úvah je zřejmé, že pokud má přechodová funkce odpovídat požadavkům na doplňování mutlimodální informace, musí mít odpovídající „syntaktické“ vlastnosti, což vede k jejímu značnému nárůstu. Nejenže musí být definovaná pro každou hodnotu argumentů v normálním režimu, ale musí být definována konsistentním způsobem i pro případ, kdy některé její hodnoty nejsou v některých stavech (např. pozorovacích) definovány. Dosáhnout toho „návrhem“ je prakticky nemožné. Jedinou schůdnou cestou nabízí proces učení, a proto se v další části soustředíme na realizaci KKA, a zejména mechanizmu autokontroly a doplňování multimodální informace, pomocí neuronových sítí.

4. REALIZACE DOPLŇOVACÍHO MECHANISMU POMOCÍ ZRCADLOVÝCH NEURONŮ

Přirozeným substrátem pro výpočetní učení jsou neuronové sítě. Uvažujme proto realizaci KKA pomocí neuronových sítí. Předpokládejme, že konečné řízení budeme realizovat pomocí neuronové sítě – to určitě lze, protože neuronové sítě jsou výpočetně ekvivalentní konečným automatům (viz např. [11]). Součásti této sítě bude i mechanizmus sebekontroly, pozůstávající

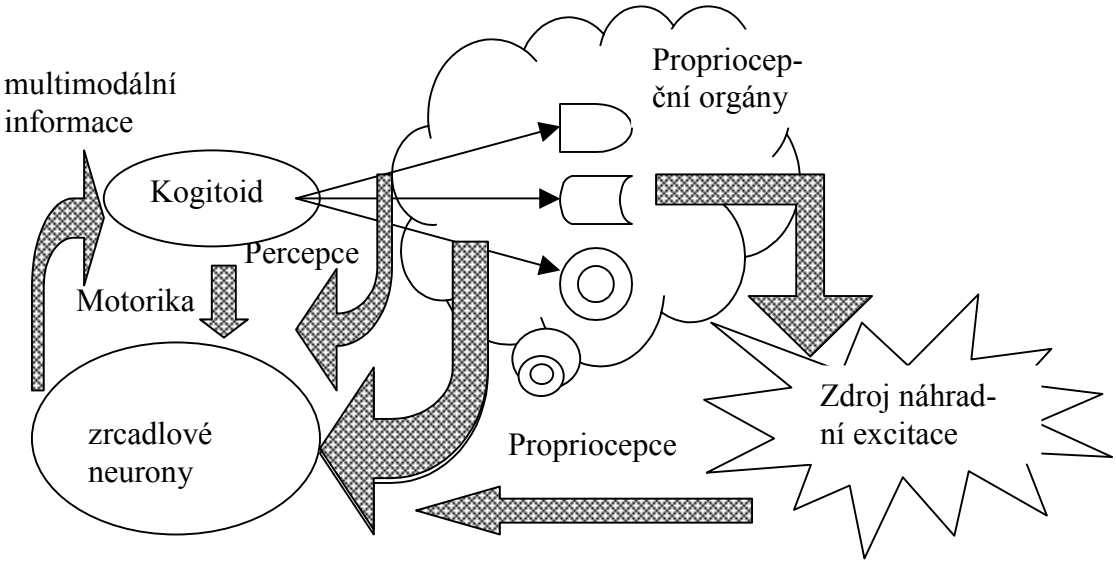
převážně ze zrcadlových neuronů. Nejprve popíšeme, jak se mechanizmus sebekontroly konfiguruje do stavu, ve kterém pracuje v normálním režimu. Dále pak ukážeme, jak lze jednoduše realizovat i mechanizmus doplňování chybějící motorické či percepční informace.

Jak již bylo řečeno, základní úlohou mechanizmu autokontroly je prověřovat, jestli argumenty přechodové funkce tvaru $Q\times(M\times S\times P)^k$ odpovídají hodnotám v normálním režimu. Tyto hodnoty se odpovídající neuronová síť učí během interakce. Nechť $(q,m_1,s_1,p_1,\ldots,m_k,s_k,p_k)$ jsou hodnoty takových argumentů pro stav q. Neuron, kontrolující současný výskyt příslušných hodnot argumentů, je zřejmě neuron, realizující booleovskou funkci AND 3k+1 příslušných proměnných v binární reprezentaci. Takový neuron – říkejme mu zrcadlový neuron -- musí existovat pro všechny možné kombinace argumentů, které jsou charakteristické pro normální režim. Namísto jednoho neuronů si můžeme představit i skupinu více neuronů, které plní daný úkol. Např. nejprve informace od senzorů prochází přes neuronový obvod, který rozezná „přípustnost“ pozorovaného pohybu, a teprve pak informace přichází do obvodu, který má na starosti posoudit „kompozici“ složené multimodální informace. Celkově se tedy rýsuje obraz neuronové sítě, do které vstupují informace o současném stavu, a pro každou SMO motorické informace vyslané řídící jednotkou pro tento SMO a senzorické a propriocepční informace od senzorů dané SMO. Tyto informace se nasměrují do všech zrcadlových neuronů paralelně a jeden z nich na ně zareaguje jako na „své“ informace. Z hlediska učení se tato síť musí naučit disjunkci konjunkcí – tj. disjunkci všech možných „smysluplných“ kombinací multimodálních argumentů a odpovídajících stavů.

Uvažujme nyní případ, že do již takto (v normálním režimu) naučené sítě se dostane neúplná či „chybná“ (tj. neodpovídající normálnímu stavu) multimodální informace – např. v režimu pozorování nesouhlasí motorická a propriocepční informace s pozorovanou senzorickou informací. Potřebujeme ale, aby zareagoval ten zrcadlový neuron, který by zareagoval, kdyby agent příslušný pohyb vykonával a pozoroval „sám na sobě“ v normálním režimu. To lze dosáhnout více způsoby. Jeden způsob je, aby se excitovala „pomocná“ neuronová síť, která do všech neuronů paralelně vyšle všechny možné naučené kombinace motorické a propriocepční informace. Pokud předpokládáme, že každá multimodální informace je jednoznačně určena svou senzorickou složkou a stavem, tak na takový „náhradní“ signál zareaguje právě ten neuron, o který nám jde. Jiná možnost je využít asociativních paměťových schopností neuronových sítí. Příslušnými detaily se zde nebudeme zabývat, nám spíše jde o to, ukázat, že takový mechanizmus existuje a lze si jeho konstrukci představit. Schéma odpovídající neuronové implementace konečného kognitivního agenta je na obr. 4.

Nyní ukážeme, že vlastnosti „našich“ zrcadlových neuronů odpovídají zrcadlovým neuronům popsaným v literatuře. Předpokládejme, že řídící jednotka systému oko—ruka vyšle v normálním režimu nějaký řídící signál své ruce i oku. Tyto signály se v příslušném zrcadlovém neuronu setkají se signálem od percepčních senzorů ruky i oka a protože jde o signály v normálním režimu, zrcadlový neuron na ně zareaguje. To znamená, že tento zrcadlový neuron je aktivní, když agent realizuje příslušný pohyb a pozoruje jej prostřednictvím vnější senzo--motorické vazby. Uvažujme nyní situaci, kdy agent A pozoruje

ruku agenta B provádějící stejný pohyb. V předchozím odstavci popsaný mechanizmus doplní tuto senzorickou informaci na úplnou multimodální informaci, na kterou opět zareaguje stejný zrcadlový neuron, jako v případě, když pohyb pozoroval agent A sám na sobě.



Obr. 4: Schéma neuronové implementace konečného kognitivního agenta

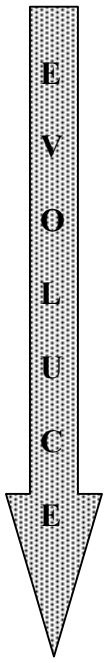
Z hlediska efektivity je možné namísto sítě pozůstávající z jednotlivých zrcadlových neuronů uvažovat pravděpodobně podstatně menší neuronovou síť, která je naučena rozeznávat potřebnou multimodální informaci a současně má i doplňovací asociační schopnosti. Všimněme si, že takovou síť můžeme považovat za implicitní realizaci světa, ve kterém se agent pohybuje a o kterém má informace. Tento svět je reprezentován množinou všech multimodálních informací o něm. Agent má k dispozici doslova jakousi mapu, která popisuje svět poznaný pomocí agentových smyslů a jeho motoriky a propriocepce. Pokud v této mapě některé části světa chybí, tak agent se s nimi ještě nesetkal, anebo prostě k nim prostřednictvím svých možností nemá přístup. K podobným závěrům o existenci implicitního modelu světa v kognitivních agentech, realizovaném pomocí neuronové sítě, dospěl i autor práce [1]. Naše úvahy jdou nad rámec úvah o významu senzomotorické informace pro kategorizaci objektů, shrnutých např. v monografii [7], protože ukazují konkrétnější mechanizmus (a sice zrcadlové neurony), na kterém může takový model být založen.

5. HYPOTETICKÝ EVOLUČNÍ ROZVOJ KOGNITIVNÍCH SCHOPNOSTÍ V KOGNITIVNÍM KONEČNÉM AGENTOVI

Uvažujme nyní hypoteticky o evolučním vývoji kognitivního agenta, který by vedl k rozvoji jeho mentálních schopností. Zde je přirozené předpokládat vývoj přes řadu vývojových

stupňů, ve kterých se budují jednotlivé mechanismy zabezpečující činnost, odpovídající současným nárokům na přežití agenta v ekologické nice, řečeno termíny vtělené kognice. Evoluci agenta musí odpovídat jeho přiměřený „tělesný“ i „duševní“ rozvoj. Oba tyto aspekty musí být v tzv. ekologické rovnováze: nemá smysl, aby agent měl PMO, které mu dodávají více informací, než je jeho mozek schopen zpracovat, anebo opačně, aby měl „výkonnější“ mozek, než stačí pro zpracování dat dodávaných jeho PMO.

Základním předpokladem kognice je tzv. „ukotvení“ (grounding) agenta v prostředí. To znamená, že agent musí vědět, jak obcovat s prostředím prostřednictví svých akcí, na základě informací, které dostává od svých percepčních orgánů. K tomu slouží mutimodální informace, uložená v jeho zrcadlové neuronové síti. Ukotvení tedy zprostředkovává a zabezpečuje jakési elementární porozumění agenta světu – na úrovni elementárních akcí agent vlastně nemůže dělat nic jiného než smysluplné akce, protože jiné se v jeho repertoáru nevyskytují, jiným se nenaučil. Pro další rozvoj agent potřebuje mít možnost imitačního učení – to zabezpečí mechanizmus autokontroly a doplňování multimodální informace. Zdá se, že tato vývojová etapa se liší od předchozí, protože existují živočichové, kteří nemají ani elementární schopnost napodobování. Schopnost napodobování přináší i emergenci konceptu „já“ (self), bez kterého nelze rozlišovat mezi akcemi, které provádí agent a pozoruje sám na sobě (tak řečeno „z vnější i vnitřní perspektivy“) , a akcemi jiných agentů.



- „Ukotvení“ elementárních akcí v kontextu a senzomotorice
- Rozlišení mezi sebou a jinými (koncept „self“), imitace elementárních akcí
- Imitace posloupností akcí pomocí učení
- Počátky komunikace prostřednictvím „body language“, pomocí „čtení mysle“ a empatie
- Přidávání vokalizace a její asociace s motorikou posunků, později s motorikou (vlastních) mluvidel
- Rozvoj „slovníku“ porozumění a „přístupu“ do něj prostřednictvím slov (jakoby nárůst počtu senzorů); vývoj odpovídající propriocepce
- **Evoluční rovnováha** mezi nárůstem množství rozlišovaných slovních podnětů a velikostí mozku
- Přímá aktivace konceptů prostřednictvím slov
- **Počátky myšlení**: nejprve hlasité „mluvení k sobě“, později převažují pohyby mluvidel, ale postupně význam pohybů klesá; příkazy k vyslovování jsou přímo asociovány s „významy“ slov
- Propriocepce aktivace abstraktních konceptů: subjektivní prožitky?
- Výsledek vývoje: myšlení jako „utlumené“ senzomotorické akce

Obr. 5: Hypotetický evoluční vývoj kognitivních schopností

Dále se přidává schopnost učení posloupnosti akcí pomocí imitace. Tuto schopnost (a další, jako učení pomocí analogie, pavlovovské reflexy, operantní podmiňování, zpracování emocí,

atd.) zabezpečuje mechanismus, označený na obr. 4 jako kogitoid. Všimněme si, že tento mechanizmus není přímo napojen na žádné periferie – je to v podstatě zařízení pro zpracováni informací, jaké se uvažovalo v rámci klasické umělé inteligence. Vstupem pro něj jsou však předzpracované, doplněné mutlimodální informace, které již zohledňují vnější svět, čím se kogitoid specializuje výlučně na zpracování „smysluplných“ percepčních informací. Kogitiod je konstruován tak, aby se učil na základě zkušeností, v různých režimech (s učitelem, bez učitele – imitací, metodou pokusu a omylů, atp.), a ve své nejrozvinutější fázi pomocí myšlení -- tj. „mentální“ simulací možných dalších scénářů dalšího vývoje situace. Režim činnosti kogitoidu závisí i na emocích. Podrobnosti o kogitoidu viz v práci [10]. S rozvojem příslušných mechanizmů učení jde ruka v ruce rozvoj komunikačních schopností a empatie. Komunikace je nejprve jednoduchá, vizuální, pomocí posunků, u některých druhů snad i pachová, a pojí se pojí s emocemi či vyhraněným kontextem. Repertoár signálů se rozšiřuje, přidávají se další komunikační kanály, zejména akustické, čím se uvolňuje zrak pro jiné důležité úkoly. Stále se buduje multimodální kombinace příslušných signálů, ve které hraje hlavní roli motorika. Dalším rozvojem klesá význam vizuální komunikace a její roli přebírá hlasová komunikace. Role motoriky se přesouvá od posunků a pohybů tělem na mluvidla. Rozvíjí se slovník pojmů, za stálého zachování jejich ukotvení v percepčně--motorických multimodálních informacích. Pojmy a dojmy lze aktivovat prostřednictvím slov, nikoliv výlučně pomocí senzorů. S aktivací pojmů možná souvisí vnitřní propriocepce této aktivity a tato propriocepce je navázaná na další multimodální informace, asociované s danými pojmy. Mluvené slovo umožňuje další strukturaci prostředí (agent rozlišuje a má k dispozici více kategorií objektů), percepční orgány jakoby začaly dodávat více rozlišitelných informací, co má za následek odpovídající rozvoj mozku. Agenti dovedou komunikovat mezi sebou a také dovedou komunikovat „sami se sebou“ prostřednictví samomluvy. Dovedou se tázat na věci, na jaké by se tázali svých „souagentů“, a sami si na své otázky odpovídají: začínají myslet. V průběhu další evoluce vokalizace při myšlení postupně ztrácí na významu, pohyby mluvil se utlumují, až převažuje pouze motorické signály (zejména pro mluvidla), které se nerealizují, ale jsou doplňované příslušnými percepčními a propriocepčními informacemi, ve kterých je tato motorika ukotvena. Agent se z hlediska své percepce tudíž nachází v prostředí jakési virtuální reality, kterou mu zprostředkují nikoliv jeho smysly, ale mechanizmy doplňování informací (které se to ovšem „naučily“ od agentových smyslů v normálním režimu). Informaci o příslušném stavu virtuální reality má agent k dispozici a možná, že zde jsou kořeny vědomí. Schematicky je rozvoj agentových kognitivních schopností naznačen na obr. 5. Naznačený vývoj mentálních schopností odpovídá i představám řady dalších autorů – za všechny jmenujme alespoň Dennetta [4], rozdíl je zejména v tom, že v našem případě jsou tyto představy podporovány daleko konkrétnějšími idejemi o realizaci mechanizmů, které podporují jednotlivé vývojové etapy evolučního rozvoje kognitivních agentů.

6. ZÁVĚR

Ke zpřesnění představ o mechanizmu myšlení v KKA vedlo několik nových původních myšlenek. První byla hypotéza, že zrcadlové neurony jsou součástí mechanizmu, který slouží pro učení, verifikaci (autokontrolu) a případné doplňování multimodální informace. Tento

mechanizmus byl zobecněn i na využívání propriocepční informace nejen jako dalšího prostředku pro ukotvení multimodální informace a zvýšení její odolnosti proti případnému výpadku některé její součásti, ale také jako základ mechanizmu vnitřního prožívání. Mechanizmus doplňováni multimodální informace dále umožnil převést zdánlivě nesouvisející percepčněmotorické operace a operace myšlení na jednotný základ. Myšlení v tomto modelu má podobu utlumených percpečněmotorických akcí, které jsou z hlediska agenta doprovázeny virtuálními prožitky a vjemy, které obstarává mechanizmus doplňování. Propriocepce tohoto stavu může být prologem ke stavu vědomí.

Naznačená teorie myšlení je vlastně teorií realizace myšlení. Popsané principy totiž byly v různé míře tušeny a předvídány mnoha filozofy, psychology a lidmi z oblasti umělé inteligence a kognitivních věd, avšak doposud chyběl jednotný výpočetní rámec, umožňující konzistentní vysvětlení mechanizmů myšlení. Zrcadlové neurony a neuronová implementace KKA navrhovaná v této práci představují možný výpočetní rámec pro teorii kognice.

LITERATURA

[1] M. Arbib: The Mirror System, Imitation, and the Evolution of Language. *Imitation in Animals and Artifacts*, Ch. Nehaniv and K. Dautenhahn, Editors, The MIT Press, to appear

[2] R. Brooks: *Cambrian Intelligence: The Early History of the New AI.* MIT Press (A Bradford Book), 1999

[3] Cruse, H.: The evolution of cognition – a hypothesis. *Cognitive Science* 27, Elsevier, 2003, s. 135-155

[4] D. C. Dennett: *Consciousness Explained*. Penguin Books, 1991, 511 p.

[5] J.R. Hurford: Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution. In: O. Kimbrough, U. Griebel, K. Plunkett (eds.): *The Evolution of Communication systems: A Comparative Approach.* The Viennna Series in Theoretical Biology, MIT Press Cambridge, MA, 2002

[6] Lengeler, J.W., Müller, B.S., di Primio, F.: Neubewertung kognitiver Leistungen im Lichte der Fähigkeiten einzelliger Lebewesen. *Kognitionswissenschaft,* 8, 2000, s. 160--178

[7] Pfeifer, R., Scheier, Ch.: *Understanding intelligence*. The MIT Press, Cambridge, Massachusetts, London, England, 1999, 697 s.

[8] V.S Ramachandran: Mirror neurons and imitation as the driving force behind "the great leap forward" in human evolution. *EDGE: The third culture*, viz http://www.edge.org/ 3rd_culture/ramachandran/ramachandran_p1.html

[9] G. Rizzolatti, L. Fadiga, V. Gallese, I. Fogassi: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3:131-141,1966

[10] J. Wiedermann, J.: Towards Algorithmic Explanation of Mind Evolution and Functioning (Invited Talk). In: *Proc. of the 23-rd International Symposium on Mathematical Foundations of Computer Science*, LNCS Vol. 1450, Springer Verlag, Berlin, 1998, pp. 152--166 .

[11] J. Wiedermann: The Computational Limits to the Cognitive Power of the Neuroidal Tabula Rasa. Journal of Experimental and Theoretical Artificial Intelligence (JETAI), v tisku, 2003